

# Conversion of SUDAAN® Output into Publication-Quality Tables—A Simplified Approach

Charlotte Cates Gard, Research Triangle Institute, Rockville, MD

## ABSTRACT

We have developed a program that utilizes the capabilities of the SAS Output Delivery System and the SAS REPORT procedure to generate Rich Text Format tables from SUDAAN PROC CROSSTAB output. Because the SAS Output Delivery System allows style elements to be specified directly in PROC REPORT, the tables generated by our program are in “final form” and require little manipulation to meet formatting requirements. In addition, as most of the formatting is done “behind the scenes,” the code is straightforward and can be easily modified for various table configurations.

## INTRODUCTION

As the survey industry expands, more programmers are required to work with data from complex surveys. Traditionally, complex survey data have been analyzed using specialized software such as SUDAAN, WesVar, and Stata. Often, the output from these programs has required further manipulation to meet the specific formatting requirements of publication.

This paper addresses the conversion of SUDAAN output into publication-quality tables. Some previous solutions have relied on the transcription of data from printed output into pre-formatted tables. Other solutions have utilized sophisticated SAS® programs, combined with manipulations in a word processing program, to generate tables. Whatever the approach, producing more than a few such tables proved to be a time-consuming and tedious task.

## THE DATA

The program presented here was used to analyze data from the 1999 National Youth Tobacco Survey. This survey, sponsored by the American Legacy Foundation, was designed to provide data on tobacco-related issues for a sample of students in grades six through twelve. More information on American Legacy Foundation research efforts can be obtained by visiting [www.americanlegacy.org](http://www.americanlegacy.org).

## THE REQUIREMENTS

A number of formatting requirements were specified for the National Youth Tobacco Survey tables. The table below depicts many of these requirements.

| Demographic Category | N    | Never-Smoker, Not Open to Smoking | Never-Smoker, Open to Smoking |
|----------------------|------|-----------------------------------|-------------------------------|
| <i>Age</i>           |      |                                   |                               |
| 11 yrs               | 1784 | 72.9<br>[69.1 - 76.8]             | 14.3<br>[12.3 - 16.3]         |
| 12 yrs               | 2477 | 57.9<br>[54.6 - 61.2]             | 16.5<br>[14.8 - 18.2]         |
| 13 yrs               | 2712 | 44.8<br>[41.4 - 48.1]             | 16.5<br>[14.5 - 18.6]         |

Here, N represents the (unweighted) sample size, summed across all values of the response variable (i.e., across the columns). For each combination of a column variable and row variable, the (weighted) row percentage and its 95% confidence interval are displayed. Those familiar with SUDAAN will recall that the CROSSTAB procedure can be used to generate sample sizes and row percentages. Confidence intervals, however, are not available in PROC CROSSTAB and must be formed within SAS.

For this project, we were required to combine the output from multiple SUDAAN cross tabulations into a single table that could be read and manipulated in Microsoft Word. Most often, cross tabulations of a response variable with the demographic variables age, race, and gender were required. Header lines were desired, to separate the data from the various cross tabulations.

## THE OLD WAY

The method initially used to generate tables for the National Youth Tobacco Survey relied on a combination of SAS programming and “by hand” manipulations in Microsoft Word. Although programmatically savvy, this approach proved to be quite labor intensive, particularly when a large number of tables were required.

The weakness of this approach lay in the hard coding of many table specifications. Changes to the specifications required that the programs be rewritten (or, at the very least, modified extensively). For instance, the number of levels of the response variable was utilized in key DATA step operations. As a result, separate programs had to be written for response variables with five levels, six levels, seven levels, etc. Because the programs relied on sophisticated programming techniques, they could be difficult for the novice programmer to understand and modify.

## THE NEW WAY

With Release 8.1, SAS introduced into production its Output Delivery System Rich Text Format capability (this capability was available experimentally in Version 8 but was significantly improved with Release 8.1). With this, files created from SAS procedural output could be read and manipulated directly in Microsoft Word.

With this in mind, we sought to develop an approach that would utilize the new Rich Text Format capability to convert SUDAAN output into publication-quality tables quickly, accurately, and with modest programming effort.

## PROC TABULATE VERSUS PROC REPORT

Originally, we intended to use the SAS TABULATE procedure to generate the required tables. However, we were unable to meet certain formatting requirements with this approach. For example, we could not create the confidence intervals in the desired format. PROC TABULATE would not allow us to display a character variable combining the (formatted) lower and upper confidence limits. We were, thus, forced to consider the lower and upper limits separately and, as a result, could not display them in the same cell, as required.

We also had problems displaying the lower and upper limits of the confidence interval below their corresponding point estimate. When we tried to do this, we ended up with a stack of cells for each row percentage—one containing the percentage itself, one containing the lower confidence limit, and one containing the upper confidence limit. In addition, with the confidence interval displayed below its corresponding point estimate, we lost the ability to display the (total) sample sizes, as with PROC TABULATE all statistics must be in the same dimension (either row, column, or page). Ultimately, the SAS REPORT procedure provided us a much better solution.

## THE DETAILS

The program presented below uses SUDAAN PROC CROSSTAB to generate (weighted) cross tabulations of age, race, and gender with smoking stage. An output data set is created by SUDAAN, which is manipulated through a series of SAS DATA step

operations. SAS PROC REPORT operates on the modified data set to generate the required table.

## SETUP

We begin our program with an OPTIONS statement. With the ORIENTATION = option, we specify the paper orientation to be used when printing to the Output Delivery System destination. Note that the OPTIONS statement can also be used to change the default paper size. This is often required when generating very large tables (tables with many columns) and can be achieved through use of the PAPERSIZE = option.

Next, formats are specified for the variables used in the cross tabulations. Formats for the demographic variables, QN1R, NEWRACE, and QN2R (which represent age, race, and gender, respectively) are as follows:

```
PROC FORMAT;
```

```
VALUE QN1RF 1 = "11 YRS"
            2 = "12 YRS"
            3 = "13 YRS"
            4 = "14 YRS"
            5 = "15 YRS"
            6 = "16 YRS"
            7 = "17 YRS"
            8 = "18-19 YRS";
```

```
VALUE NEWRACEF 1 = "WHITES"
               2 = "AFR-AMER"
               3 = "HISPANICS"
               4 = "ASIANS"
               5 = "OTHERS";
```

```
VALUE QN2RF 1 = "MALES"
            2 = "FEMALES";
```

The format for the response variable, STAGE, is also provided. Note the use of the "\ " character in certain format values. This is called a split character and allows us to wrap these values for display in the final table.

```
VALUE STAGEF 1 = "NEVER-SMOKER,\NOT OPEN\TO
                 SMOKING"
              2 = "NEVER-SMOKER,\OPEN TO\SMOKING"
              3 = "EXPERIMENTER"
              4 = "FORMER SMOKER"
              5 = "NON-DAILY\CURRENT SMOKER"
              6 = "ESTABLISHED\SMOKER";
```

During setup, each of the above variables is created in the SUSC data set and the proper format is applied. (Note that LIBNAME statements, used to indicate storage locations for the input and output files, should also be specified during setup. They have been omitted here.)

## MACRO VARIABLES

A number of macro variables are created, using the %LET statement, as below.

```
%LET COL_VAR = STAGE;
%LET LEVELS = 6;
```

COL\_VAR specifies the response variable (here STAGE), and LEVELS specifies the number of levels of the response variable (here six). Although not shown here, macro variables are also used to specify the macro variables FILENAME and TITLE.

Note, with this particular approach, the %LET statement is the only place where the number of levels of the response variable is indicated. Otherwise, this program does not need to know the number of levels of the response variable. Therefore, if the response variable was redefined, resulting in a change in the number of levels, this is the only line of code that would have to

be modified. Furthermore, if one wished to represent cross tabulations of age, race, and gender with another response variable (brand preference, for instance), he or she could do so simply by changing these %LET statements.

## SUDAAN PROC CROSSTAB

We are now ready to run the SUDAAN CROSSTAB procedure. (Recall that, before using a SUDAAN procedure, we must sort the data set by the variables appearing on the NEST statement. This step is omitted here.)

```
PROC CROSSTAB DATA = YTS.SUSC FILETYPE = SAS
              DESIGN = WR;
  NEST NSTRATUM NPSU / MISSUNIT;
  WEIGHT FINALWT;
  SUBGROUP QN1R NEWRACE QN2R &COL_VAR;
  LEVELS 8 5 2 &LEVELS;
  TABLES (QN1R NEWRACE QN2R) * &COL_VAR;
  PRINT NSUM ROWPER SEROW;
  OUTPUT NSUM ROWPER SEROW /
         FILENAME =OUTPUT.SUDOUT
         FILETYPE = SAS REPLACE;
RUN;
```

With the PRINT and OUTPUT statements, we request that the following statistics be output: NSUM (the unweighted sample size in each cell), ROWPER (the row percentage in each cell), and SEROW (the standard error of the row percentage in each cell). The PRINT statement requests that this information be printed to the screen, whereas, the OUTPUT statement requests that this information be written to a file (specifically to the file SUDOUT in the path specified by the LIBNAME statement for OUTPUT).

Note that PROC CROSSTAB generates a separate table for each cross tabulation requested. SUDAAN numbers the tables using the variable TABLENO. Here, TABLENO = 1 is the table for QN1R crossed with COL\_VAR, TABLENO = 2 is the table for NEWRACE crossed with COL\_VAR, and TABLENO = 3 is the table for QN2R crossed with COL\_VAR.

A subset of the SUDOUT data set is provided below.

| PROCNUM | TABLENO | QN1R | STAGE | NEWRACE | QN2R | NSUM  | ROWPER   | SEROW    |
|---------|---------|------|-------|---------|------|-------|----------|----------|
| 1       | 1       | 0    | 0     | -2      | -2   | 14589 | 100      | 0        |
| 1       | 1       | 0    | 1     | -2      | -2   | 6024  | 40.29219 | 1.544229 |
| 1       | 1       | 0    | 2     | -2      | -2   | 1744  | 11.75898 | 0.560587 |
| 1       | 1       | 0    | 3     | -2      | -2   | 5278  | 36.50132 | 1.247278 |
| 1       | 1       | 0    | 4     | -2      | -2   | 173   | 1.314009 | 0.153495 |
| 1       | 1       | 0    | 5     | -2      | -2   | 388   | 2.860266 | 0.262266 |
| 1       | 1       | 0    | 6     | -2      | -2   | 982   | 7.273234 | 0.694255 |
| 1       | 1       | 1    | 0     | -2      | -2   | 1784  | 100      | 0        |
| 1       | 1       | 1    | 1     | -2      | -2   | 1310  | 72.93052 | 1.95798  |
| 1       | 1       | 1    | 2     | -2      | -2   | 248   | 14.30842 | 1.021948 |
| 1       | 1       | 1    | 3     | -2      | -2   | 223   | 12.5744  | 1.699181 |
| 1       | 1       | 1    | 4     | -2      | -2   | 1     | 0.051571 | 0.052627 |
| 1       | 1       | 1    | 5     | -2      | -2   | 1     | 0.070561 | 0.06979  |
| 1       | 1       | 1    | 6     | -2      | -2   | 1     | 0.064537 | 0.063832 |

## DATA STEP MANIPULATIONS

A number of manipulations are required, to ready the SUDAAN output data set for SAS PROC REPORT.

```
DATA DATA1;
  SET OUTPUT.SUDOUT;
  IF &COL_VAR = 0 THEN DELETE;
  GRP = TABLENO;
  LCL = ROUND(ROWPER - (1.96 * SEROW), .1);
  IF LCL < 0 THEN LCL = 0;
  UCL = ROUND(ROWPER + (1.96 * SEROW), .1);
  LENGTH CI $15 VALUE $51;
  CI = "[" || COMPRESS(PUT(LCL, 5.1)) ||
        " - " || COMPRESS(PUT(UCL, 5.1)) || "]";
  VALUE = PUT(ROUND(ROWPER, .1), 5.1) ||
        " " || CI;
RUN;
```

In the code above, we create a temporary data set, DATA1, from the SUDAAN output data set. We remove observations for which the macro variable COL\_VAR is equal to zero. These observations represent the totals, summed across the columns. Although these totals are included in the final table, we will use PROC REPORT to recalculate them later.

Next, we create the variable GRP, which is equal to TABLENO. LCL and UCL, the (rounded) lower and upper confidence limits, are then calculated.

The character variable CI is formed by concatenating LCL and UCL (with leading and trailing brackets and an intervening dash). Similarly, the variable VALUE, is formed by concatenating the (rounded) value of ROWPER and the formatted confidence interval, CI.

```
DATA DATA2;
SET DATA1;
IF (QN1R = 0 or NEWRACE = 0 or QN2R = 0)
    THEN DO;
NSUM = .;
VALUE = "          ";
END;
RUN;
```

Here, we create a data set, DATA2, from DATA1. We identify those observations for which QN1R is equal to zero, NEWRACE is equal to zero, or QN2R is equal to zero. These represent the overall totals for QN1R, NEWRACE, and QN2R (summed across the rows). By setting NSUM and VALUE equal to missing (or blank) for these observations, we generate header rows (blank lines that appear before the age, race, and gender data in the final table).

```
DATA DATA3;
SET DATA1;
IF QN1R = 0 THEN DO;
GRP = 4;
NSUM = .;
VALUE = "          ";
OUTPUT;
END;
RUN;
```

Next, we create the data set DATA3, also from DATA1. We create a fourth group from those observations for which QN1R is equal to zero. This serves to add a blank line before the overall totals. (For convenience, we use observations for which QN1R is equal to zero. We might also have used observations for which NEWRACE is equal to zero or for which QN2R is equal to zero. The key is to have one observation for each (nonzero) value of COL\_VAR.)

```
DATA DATA4;
SET DATA1;
IF QN1R = 0 THEN DO;
GRP = 5;
OUTPUT;
END;
RUN;
```

DATA4 is created from the data set DATA1 and contains the totals based on age. These totals will be used as overall totals in the final table.

```
DATA DATA5;
SET DATA2 DATA3 DATA4;
RUN;
```

Here, we merge the data sets we have created above. You will recall that DATA2 contains the age data, race data, and gender data, DATA3 contains a blank line (effectively), and DATA4 contains the overall (age) totals.

```
DATA DATA6;
SET DATA5;
IF GRP = 1 THEN TEMP = QN1R;
ELSE IF GRP = 2 THEN TEMP = NEWRACE;
ELSE IF GRP = 3 THEN TEMP = QN2R;
ELSE IF GRP = 4 THEN TEMP = QN1R;
ELSE IF GRP = 5 THEN TEMP = QN1R;
RUN;
```

With DATA6, we create the variable TEMP. Note that, for each GRP, TEMP is set equal to the variable from which the data were derived. (Recall that GRP = 1 is derived from TABLENO = 1, which is derived from the QN1R cross tabulation. GRP = 2 is derived from the NEWRACE cross tabulation, and GRP = 3 is derived from the QN2R cross tabulation. GRP = 4 and GRP = 5 are derived from QN1R.)

```
DATA DATA7;
SET DATA6;
BY GRP TEMP;
RETAIN LEVEL 0;
IF FIRST.TEMP THEN DO;
LEVEL = LEVEL + 1;
END;
RUN;
```

The GRP variable is used to distinguish between age data, race data, gender data, the blank line preceding the age totals, and the overall (age) totals. With this DATA step, LEVEL is incremented each time the value of TEMP changes. This effectively numbers the rows in the final table.

### FORMATS

We now specify the format for the newly created variable LEVEL. To do this, we use the FORMAT procedure in SAS. Note that we have included here not only the values of each of the variables (QN1R, NEWRACE, and QN2R), but also headers for each of these variables.

```
PROC FORMAT;

VALUE LEVELF 1 = "AGE"
              2 = "11 YRS"
              3 = "12 YRS"
              4 = "13 YRS"
              5 = "14 YRS"
              6 = "15 YRS"
              7 = "16 YRS"
              8 = "17 YRS"
              9 = "18-19 YRS"
              10 = "RACE"
              11 = "WHITES"
              12 = "AFR-AMER"
              13 = "HISPANICS"
              14 = "ASIANS"
              15 = "OTHERS"
              16 = "GENDER"
              17 = "MALES"
              18 = "FEMALES"
              19 = " "
              20 = "TOTAL";
```

```
VALUE MISSF . = "          ";
```

```
RUN;
```

One disadvantage of this approach is that the rows of the table must be specified in the SAS program (and, thus, must be known in advance). This is necessary because we are combining cross tabulations for age, race, and gender in one table. If only one cross tabulation were required, this would not be necessary (and the number of DATA step manipulations would be greatly reduced).

## STYLE DEFINITION

Next, we use the SAS TEMPLATE procedure to create a style definition. The SAS System provides a number of style definitions for use in formatting Output Delivery System output. These style definitions can be used as provided or can be modified to add new style elements or change existing style elements.

```
PROC TEMPLATE;
DEFINE STYLE STYLES.LEGACY;
PARENT = STYLES.RTF;
STYLE HEADER FROM HEADERSANDFOOTERS /
    BACKGROUND = WHITE;
STYLE SYSTEMTITLE FROM TITLEANDFOOTERS /
    FONT = ("TIMES", 10PT);
STYLE SYSTEMFOOTER FROM TITLEANDFOOTERS /
    FONT = ("TIMES", 10PT);
END;
RUN;
```

In the above code, we use PROC TEMPLATE to modify an existing style definition--STYLES.RTF. With the STYLE HEADER statement, we change the background color for the header from grey to white. With the STYLE SYSTEMTITLE and STYLE SYSTEMFOOTER statements, we specify the font type and size for the title and footnote.

The following code applies the style definition:

```
ODS LISTING CLOSE;
ODS RTF FILE = &FILENAME STYLE = LEGACY;
```

This code also tells the SAS System to close the Listing destination and to begin writing output to the Rich Text Format file specified by the macro variable FILENAME.

## THE REPORT PROCEDURE

We are now ready to run PROC REPORT. PROC REPORT operates on DATA7, the final manipulated data set, to generate the formatted table.

```
PROC REPORT DATA = DATA7 SPLIT = "\" NOWD
    STYLE(COLUMN) = [FONT_SIZE = 8PT JUST = C];
COLUMNS GRP LEVEL NSUM &COL_VAR,VALUE N;
DEFINE GRP / GROUP NOPRINT;
DEFINE LEVEL / GROUP ORDER = INTERNAL
    FORMAT = LEVELF. " ";
    STYLE = {CELLWIDTH = .75IN};
DEFINE NSUM / ANALYSIS FORMAT = MISSF. " "
    STYLE = {CELLWIDTH = .75IN};
DEFINE &COL_VAR / ACROSS ORDER = INTERNAL " "
    STYLE = {CELLWIDTH = 1IN};
DEFINE VALUE / DISPLAY " "
    STYLE = {CELLWIDTH = 1IN} FLOW;
DEFINE N / NOPRINT;
TITLE1 &TITLE;
FOOTNOTE "**DATA ARE FROM 1999 NATIONAL YOUTH
    TOBACCO SURVEY.";
RUN;
```

The NOWD option on the PROC statement runs PROC REPORT without the REPORT window and is required when using the Output Delivery System with PROC REPORT.

Recall that VALUE is a character variable combining each row percent with its corresponding (formatted) confidence interval. With &COL\_VAR,VALUE in the COLUMNS statement, we are requesting that VALUE be displayed, for each level of COL\_VAR, for each GRP and LEVEL combination.

The use of FLOW in the DEFINE statement for VALUE allows this particular column to flow to another line if it exceeds a specified width. This solves the problem of displaying the confidence interval below its corresponding point estimate as we have set our

variable lengths and column widths to ensure that this flow takes place.

N, here, is a dummy variable. This program will not work without it. As you can see in the DEFINE statement for N, it is not actually printed.

Note that, in addition to the style elements specified in the style definition, we can also specify style elements directly in PROC REPORT. Here, we have specified the font size, the justification of text in the columns (centered), and the cell widths.

We must now tell the SAS System that we wish to stop writing the output to the Rich Text Format file. We do so with the following code:

```
ODS RTF CLOSE;
ODS LISTING;
```

This closes the Rich Text Format destination and reopens the Listing destination.

Appendix A displays the table that is generated by PROC REPORT.

## "BY HAND" MANIPULATIONS

Although the table generated by our program satisfies the key formatting requirements, certain minor "by hand" manipulations must still be made. These include the following: the addition of a superscript character in the title and footnote, the addition of (one) gridline and (two) labels in the topmost header row, the application of italics to certain labels, and the application of shading to the header rows. In some cases, margins must also be adjusted.

Provided in Appendix B is the table generated by our program, modified slightly for publication.

## CONCLUSION

With the introduction of the SAS Output Delivery System Rich Text Format capability, the opportunity exists to simplify the process of converting SUDAAN output into publication-quality tables. The program described above offers one approach for doing so. With it, we have realized significant reductions in the amount of time required to generate publication-quality tables for the National Youth Tobacco Survey.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Charlotte Gard  
Research Triangle Institute  
6110 Executive Boulevard, Suite 420  
Rockville, MD 20852  
Work Phone: (301) 770-8226  
Fax: (301) 230-4646  
E-mail: [cgard@rti.org](mailto:cgard@rti.org)

SAS is a registered trademark or trademark of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

SUDAAN® is a trademark of the Research Triangle Institute.

Other brand and product names are registered trademarks or trademarks of their respective companies.

APPENDIX A: Table A-2. Percentage Distribution of Smoking Stage\*

|           |       | Never-Smoker,<br>Not Open<br>to Smoking | Never-Smoker,<br>Open to<br>Smoking | Experimenter          | Former Smoker      | Non-Daily<br>Current Smoker | Established<br>Smoker |
|-----------|-------|-----------------------------------------|-------------------------------------|-----------------------|--------------------|-----------------------------|-----------------------|
| Age       |       |                                         |                                     |                       |                    |                             |                       |
| 11 yrs    | 1784  | 72.9<br>[69.1 - 76.8]                   | 14.3<br>[12.3 - 16.3]               | 12.6<br>[9.2 - 15.9]  | 0.1<br>[0.0 - 0.2] | 0.1<br>[0.0 - 0.2]          | 0.1<br>[0.0 - 0.2]    |
| 12 yrs    | 2477  | 57.9<br>[54.6 - 61.2]                   | 16.5<br>[14.8 - 18.2]               | 24.6<br>[21.3 - 28.0] | 0.1<br>[0.0 - 0.1] | 0.3<br>[0.1 - 0.5]          | 0.6<br>[0.3 - 0.9]    |
| 13 yrs    | 2712  | 44.8<br>[41.4 - 48.1]                   | 16.5<br>[14.5 - 18.6]               | 34.7<br>[31.6 - 37.9] | 0.3<br>[0.1 - 0.5] | 1.4<br>[0.8 - 2.1]          | 2.2<br>[1.5 - 3.0]    |
| 14 yrs    | 1895  | 33.0<br>[30.2 - 35.9]                   | 14.2<br>[12.0 - 16.4]               | 43.4<br>[39.1 - 47.8] | 1.0<br>[0.4 - 1.5] | 2.8<br>[1.9 - 3.8]          | 5.6<br>[4.0 - 7.1]    |
| 15 yrs    | 1747  | 31.4<br>[27.6 - 35.3]                   | 8.9<br>[7.4 - 10.4]                 | 43.6<br>[40.2 - 46.9] | 1.9<br>[1.1 - 2.7] | 4.2<br>[3.0 - 5.4]          | 10.0<br>[7.8 - 12.1]  |
| 16 yrs    | 1761  | 24.8<br>[21.9 - 27.7]                   | 6.7<br>[5.6 - 7.7]                  | 47.9<br>[44.8 - 51.1] | 3.0<br>[2.1 - 3.9] | 4.8<br>[3.4 - 6.1]          | 12.8<br>[10.1 - 15.5] |
| 17 yrs    | 1734  | 24.7<br>[21.6 - 27.7]                   | 5.2<br>[4.0 - 6.5]                  | 44.5<br>[41.0 - 47.9] | 2.3<br>[1.6 - 3.0] | 5.6<br>[4.1 - 7.1]          | 17.8<br>[14.0 - 21.5] |
| 18-19 yrs | 479   | 19.0<br>[14.2 - 23.8]                   | 5.9<br>[3.4 - 8.3]                  | 42.1<br>[37.4 - 46.9] | 4.6<br>[2.5 - 6.6] | 7.0<br>[5.1 - 9.0]          | 21.4<br>[15.6 - 27.2] |
| Race      |       |                                         |                                     |                       |                    |                             |                       |
| Whites    | 8345  | 40.6<br>[36.6 - 44.5]                   | 11.7<br>[10.2 - 13.2]               | 32.9<br>[30.1 - 35.8] | 1.6<br>[1.1 - 2.0] | 3.7<br>[3.0 - 4.3]          | 9.5<br>[7.6 - 11.5]   |
| Afr-Amers | 2336  | 39.5<br>[35.5 - 43.4]                   | 11.4<br>[9.9 - 12.9]                | 44.9<br>[41.2 - 48.6] | 0.9<br>[0.4 - 1.4] | 0.8<br>[0.4 - 1.2]          | 2.6<br>[1.6 - 3.5]    |
| Hispanics | 2623  | 39.0<br>[35.2 - 42.7]                   | 12.4<br>[10.7 - 14.2]               | 42.0<br>[38.2 - 45.8] | 0.9<br>[0.4 - 1.3] | 1.9<br>[1.2 - 2.6]          | 3.8<br>[2.5 - 5.0]    |
| Asians    | 538   | 54.4<br>[48.3 - 60.5]                   | 11.6<br>[8.6 - 14.7]                | 25.6<br>[20.6 - 30.6] | 1.3<br>[0.3 - 2.4] | 1.3<br>[0.0 - 2.7]          | 5.8<br>[2.5 - 9.2]    |
| Others    | 455   | 33.8<br>[27.4 - 40.2]                   | 13.7<br>[8.7 - 18.8]                | 37.6<br>[31.8 - 43.3] | 1.3<br>[0.1 - 2.4] | 4.6<br>[1.4 - 7.8]          | 9.0<br>[4.9 - 13.2]   |
| Gender    |       |                                         |                                     |                       |                    |                             |                       |
| Males     | 7265  | 39.0<br>[36.1 - 41.9]                   | 11.7<br>[10.3 - 13.1]               | 36.5<br>[34.1 - 38.8] | 1.7<br>[1.3 - 2.2] | 3.3<br>[2.6 - 4.1]          | 7.7<br>[6.3 - 9.2]    |
| Females   | 7267  | 41.5<br>[38.0 - 45.1]                   | 11.8<br>[10.7 - 12.9]               | 36.6<br>[33.7 - 39.5] | 0.9<br>[0.6 - 1.2] | 2.4<br>[1.9 - 2.9]          | 6.8<br>[5.4 - 8.2]    |
| Total     | 14589 | 40.3<br>[37.3 - 43.3]                   | 11.8<br>[10.7 - 12.9]               | 36.5<br>[34.1 - 38.9] | 1.3<br>[1.0 - 1.6] | 2.9<br>[2.3 - 3.4]          | 7.3<br>[5.9 - 8.6]    |

\*Data are from 1999 National Youth Tobacco Survey.

APPENDIX B: Table A-2. Percentage Distribution of Smoking Stage<sup>1</sup> (MODIFIED)

| Demographic Category | N     | Never-Smoker, Not Open to Smoking | Never-Smoker, Open to Smoking | Experimenter          | Former Smoker      | Non-Daily Current Smoker | Established Smoker    |
|----------------------|-------|-----------------------------------|-------------------------------|-----------------------|--------------------|--------------------------|-----------------------|
| <i>Age</i>           |       |                                   |                               |                       |                    |                          |                       |
| 11 yrs               | 1784  | 72.9<br>[69.1 - 76.8]             | 14.3<br>[12.3 - 16.3]         | 12.6<br>[9.2 - 15.9]  | 0.1<br>[0.0 - 0.2] | 0.1<br>[0.0 - 0.2]       | 0.1<br>[0.0 - 0.2]    |
| 12 yrs               | 2477  | 57.9<br>[54.6 - 61.2]             | 16.5<br>[14.8 - 18.2]         | 24.6<br>[21.3 - 28.0] | 0.1<br>[0.0 - 0.1] | 0.3<br>[0.1 - 0.5]       | 0.6<br>[0.3 - 0.9]    |
| 13 yrs               | 2712  | 44.8<br>[41.4 - 48.1]             | 16.5<br>[14.5 - 18.6]         | 34.7<br>[31.6 - 37.9] | 0.3<br>[0.1 - 0.5] | 1.4<br>[0.8 - 2.1]       | 2.2<br>[1.5 - 3.0]    |
| 14 yrs               | 1895  | 33.0<br>[30.2 - 35.9]             | 14.2<br>[12.0 - 16.4]         | 43.4<br>[39.1 - 47.8] | 1.0<br>[0.4 - 1.5] | 2.8<br>[1.9 - 3.8]       | 5.6<br>[4.0 - 7.1]    |
| 15 yrs               | 1747  | 31.4<br>[27.6 - 35.3]             | 8.9<br>[7.4 - 10.4]           | 43.6<br>[40.2 - 46.9] | 1.9<br>[1.1 - 2.7] | 4.2<br>[3.0 - 5.4]       | 10.0<br>[7.8 - 12.1]  |
| 16 yrs               | 1761  | 24.8<br>[21.9 - 27.7]             | 6.7<br>[5.6 - 7.7]            | 47.9<br>[44.8 - 51.1] | 3.0<br>[2.1 - 3.9] | 4.8<br>[3.4 - 6.1]       | 12.8<br>[10.1 - 15.5] |
| 17 yrs               | 1734  | 24.7<br>[21.6 - 27.7]             | 5.2<br>[4.0 - 6.5]            | 44.5<br>[41.0 - 47.9] | 2.3<br>[1.6 - 3.0] | 5.6<br>[4.1 - 7.1]       | 17.8<br>[14.0 - 21.5] |
| 18-19 yrs            | 479   | 19.0<br>[14.2 - 23.8]             | 5.9<br>[3.4 - 8.3]            | 42.1<br>[37.4 - 46.9] | 4.6<br>[2.5 - 6.6] | 7.0<br>[5.1 - 9.0]       | 21.4<br>[15.6 - 27.2] |
| <i>Race</i>          |       |                                   |                               |                       |                    |                          |                       |
| Whites               | 8345  | 40.6<br>[36.6 - 44.5]             | 11.7<br>[10.2 - 13.2]         | 32.9<br>[30.1 - 35.8] | 1.6<br>[1.1 - 2.0] | 3.7<br>[3.0 - 4.3]       | 9.5<br>[7.6 - 11.5]   |
| Afr-Amer             | 2336  | 39.5<br>[35.5 - 43.4]             | 11.4<br>[9.9 - 12.9]          | 44.9<br>[41.2 - 48.6] | 0.9<br>[0.4 - 1.4] | 0.8<br>[0.4 - 1.2]       | 2.6<br>[1.6 - 3.5]    |
| Hispanics            | 2623  | 39.0<br>[35.2 - 42.7]             | 12.4<br>[10.7 - 14.2]         | 42.0<br>[38.2 - 45.8] | 0.9<br>[0.4 - 1.3] | 1.9<br>[1.2 - 2.6]       | 3.8<br>[2.5 - 5.0]    |
| Asians               | 538   | 54.4<br>[48.3 - 60.5]             | 11.6<br>[8.6 - 14.7]          | 25.6<br>[20.6 - 30.6] | 1.3<br>[0.3 - 2.4] | 1.3<br>[0.0 - 2.7]       | 5.8<br>[2.5 - 9.2]    |
| Others               | 455   | 33.8<br>[27.4 - 40.2]             | 13.7<br>[8.7 - 18.8]          | 37.6<br>[31.8 - 43.3] | 1.3<br>[0.1 - 2.4] | 4.6<br>[1.4 - 7.8]       | 9.0<br>[4.9 - 13.2]   |
| <i>Gender</i>        |       |                                   |                               |                       |                    |                          |                       |
| Males                | 7265  | 39.0<br>[36.1 - 41.9]             | 11.7<br>[10.3 - 13.1]         | 36.5<br>[34.1 - 38.8] | 1.7<br>[1.3 - 2.2] | 3.3<br>[2.6 - 4.1]       | 7.7<br>[6.3 - 9.2]    |
| Females              | 7267  | 41.5<br>[38.0 - 45.1]             | 11.8<br>[10.7 - 12.9]         | 36.6<br>[33.7 - 39.5] | 0.9<br>[0.6 - 1.2] | 2.4<br>[1.9 - 2.9]       | 6.8<br>[5.4 - 8.2]    |
| Total                | 14589 | 40.3<br>[37.3 - 43.3]             | 11.8<br>[10.7 - 12.9]         | 36.5<br>[34.1 - 38.9] | 1.3<br>[1.0 - 1.6] | 2.9<br>[2.3 - 3.4]       | 7.3<br>[5.9 - 8.6]    |

<sup>1</sup>Data are from 1999 National Youth Tobacco Survey.